ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE AND MANAGEMENT







УДК 004.89
https://doi.org/10.23947/2687-1653-2023-23-1-66-75

Научная статья



Модель машинного обучения для обнаружения COVID-19 на ранней стадии по аномалиям в ритме сердца

М.С. Межов[1] ✉, **В.О. Козицин**[1], **Ю.Д. Кацер**[2]

[1]ООО «Цифровые технологии и платформы», Российская Федерация, Москва, ул. Дубининская, 53, стр. 6

[2]Сколковский институт науки и технологии, Российская Федерация, Москва, территория инновационного центра «Сколково», Большой бульвар, 30, стр. 1

✉ msmezhov@ya.ru

Аннотация

Введение. Электронные устройства, способные собирать данные по телеметрии индивидуума, открыли перспективы доклинического выявления признаков COVID-19. Известные решения предполагают анализ информации, которую сложно получить в моменте. Речь идет, например, о состоянии крови или ПЦР-тесте. Это существенно ограничивает возможности интеграции алгоритмов с наручными гаджетами. При этом сердечно-сосудистая система как объект наблюдения достаточно информативна, съем данных хорошо проработан. В статье описана задача детекции ковидных аномалий в ритмограммах. Цель работы — создание математической модели на базе алгоритмов машинного обучения для автоматизации процесса выявления ковидных аномалий в ритме сердца. Показана возможность интеграции полученных результатов с фитнесс-браслетами и умными часами.

Материалы и методы. В работе задействовали открытый стек технологий: Python, Scikit-learn, Lightgbm. При оценке качества моделей для бинарной классификации использовалась метрика F_1. Изучены 229 ритмограмм сердца (кардиоинтервалографий) пациентов с COVID-19. Наличие или отсутствие признаков аномалии определялось с учетом времени ритмограммы и интервалов между сердцебиениями. Графически показаны отклонения, которые могут свидетельствовать о заражении. По итогам разведочного анализа собран перечень признаков, указывающих на аномалию.

Результаты исследования. В результате проделанной работы получена математическая модель, которая детектирует специфичные для COVID-19 аномалии сердечного ритма с точностью 83 %. Выявлены и ранжированы основные признаки, определяющие прогностическую способность модели. Это текущее значение интервала между ударами сердца, производные в последующей и предыдущей точках измерения продолжительности сердцебиения, первая производная в текущей точке и отклонение от медианы текущего значения длительности *RR*-интервала. Первый показатель в этом перечне признан наиболее значимым, последний — наименее. Для целей машинного обучения оценивался потенциал пяти алгоритмов: IsolationForest, LGBMClassifier, RandomForestClassifier, ExtraTreesClassifier, SGDOneClassSVM. Визуализированы нормальные и аномальные результаты наблюдений в изолирующих деревьях. Установлен параметр, который соответствует вероятности регулярного наблюдения за пределами нормы, и выбрано его значение — 0,11. С учетом данного показателя построен график для модели SGDOneClassSVM. По набору данных с применением техники перекрестной проверки рассчитана метрика качества. Речь идет о ритмограмме с временны́м рядом наблюдений, снятых за один непрерывный интервал времени у одного человека. Описан пошаговый процесс получения усредненных значений метрики для каждой модели. При сравнении самый высокий показатель зафиксирован у модели LGBMClassifier, наименьшие — у SGDOneClassSVM и IsolationForest.

Обсуждение и заключения. Полученная математическая модель занимает мало места в памяти мобильного устройства, то есть не предъявляет значимых требований к вычислительным ресурсам. Решение обладает

приемлемым качеством детекции для доклинического скрининга связанных с COVID-19 сердечно-сосудистых нарушений. Алгоритм обнаруживает аномалии в 83 % случаев. Для записи ритмограммы достаточно 4 минут. Предлагаемый сценарий использования интегрированного решения лаконичен и легко реализуем. Широкое использование разработки может способствовать выявлению COVID-19 на ранней стадии.

Ключевые слова: COVID-19, причины смерти ковид-положительных пациентов, осложнения в работе сердечно-сосудистой системы, ПЦР-тест, доклинический контроль сердечно-сосудистой системы, встроенные датчики частоты пульса, ритмограмма, RR-интервал, электрокардиограмма сердца, аномальное по продолжительности сердцебиение, сердцебиение с аномальным ритмом, машинное обучение, алгоритм LGBMClassifier.

Благодарности. Авторы выражают благодарность руководству и модераторам открытого всероссийского соревнования профессионалов в сфере цифровой экономики «Цифровой прорыв» за предоставленные данные для исследования.

Для цитирования. Межов М.С., Козицин В.О., Кацер Ю.Д. Модель машинного обучения для обнаружения COVID-19 на ранней стадии по аномалиям в ритме сердца. *Advanced Engineering Research (Rostov-on-Don).* 2023;23(1):66–75. https://doi.org/10.23947/2687-1653-2023-23-1-66-75

Original article

Machine Learning Model for Early Detection of COVID-19 by Heart Rhythm Abnormalities

Maksim S Mezhov[1] ✉, **Vyacheslav O Kozitsin**[1], **Iurii D Katser**[2]

[1] “Digital Technologies and Platforms” LLC, 53, Dubininskaya St., Moscow, Russian Federation

[2] Skolkovo Institute of Science and Technology, 30, Bolshoy Boulevard, Moscow, Russian Federation

✉ msmezhov@ya.ru

Abstract

Introduction. Electronic devices capable of collecting individual telemetry data have opened up prospects for preclinical detection of COVID-19 signs. Known solutions involve the analysis of information that is difficult to obtain at the moment. We are talking, specifically, about the blood condition or a PCR test. This significantly limits the possibility of integrating algorithms with wrist gadgets. At the same time, the cardiovascular system as an object of observation is quite informative, the data collection is well developed. The article describes the problem of detecting covid anomalies in rhythm strips. The work aims at creating a mathematical model based on machine learning algorithms to automate the process of detecting covid abnormalities in the heart rhythm. The possibility of integrating the results obtained with fitness bracelets and smart watches is shown.

Materials and Methods. The work involved an open technology stack: Python, Scikit-learn, Lightgbm. When assessing the quality of models for binary classification, metric F_1 was used. 229 cardiac rhythm strips (cardiointervalographies) of patients with COVID-19 were studied. The presence or absence of signs of an anomaly was determined taking into account the time of the rhythm strip and the intervals between heartbeats. Deviations that could indicate infection were shown graphically. Based on the exploratory analysis results, a list of signs indicating an anomaly was made.

Results. As a result of the work done, a mathematical model was obtained that detected heart rate abnormalities specific to COVID-19 with an accuracy of 83 %. The basic features determining the predictive ability of the model were identified and ranked. They included the current value of the interval between heartbeats, the derivatives at the subsequent and previous points of measuring the duration of the heartbeat, the first derivative at the current point, and the deviation of the current value of the duration of the *RR*-interval from the median. The first indicator in this list was recognized as the most significant, the last — the least. For machine learning purposes, the potential of five algorithms was evaluated: IsolationForest, LGBMClassifier, RandomForestClassifier, ExtraTreesClassifier, SGDOneClassSVM. The normal and abnormal results of observations in isolation trees were visualized. A parameter was set that corresponded to the probability of regular observation outside the norm, and its value was selected — 0.11. Taking into account this indicator, a graph was constructed for the SGDOneClassSVM model. Based on the data set, using the cross-validation technique, the quality metric was calculated. The case in hand was a rhythm strip with a time series of observations taken in one continuous time interval from one person. A step-by-step process of obtaining averaged metric values for each model was described. In comparison, the highest indicator was recorded for the LGBMClassifier model, the lowest — for SGDOneClassSVM and IsolationForest.

Discussion and Conclusions. The resulting mathematical model takes up little space in the memory of a mobile device, i.e., it does not impose significant requirements on computing resources. The solution has an acceptable detection quality for preclinical screening of COVID-19-related cardiovascular disorders. The algorithm detects anomalies in 83 % of cases. Four minutes is enough to record a rhythm strip. The proposed scenario for using an integrated solution is concise and easy to implement. Widespread use of the development can contribute to the detection of COVID-19 at an early stage.

Keywords: COVID-19, causes of death in covid-positive patients, complications in the work of cardiovascular system, PCR test, preclinical monitoring of the cardiovascular system, built-in pulse rate sensors, rhythm strip, RR-interval, cardiac electrocardiogram, abnormal heartbeat, heartbeat with abnormal rhythm, machine learning, LGBMClassifier algorithm.

Acknowledgements. The authors would like to thank the management and moderators of the open All-Russian competition of professionals in the digital economy “Digital Breakthrough” for the data provided for the study.

For citation. Mezhov MS, Kozitsin VO, Katser ID. Machine Learning Model for Early Detection of COVID-19 by Heart Rhythm Abnormalities. *Advanced Engineering Research (Rostov-on-Don).* 2023;23(1):66–75. https://doi.org/10.23947/2687-1653-2023-23-1-66-75

Введение. Изучение влияния COVID-19 на человека остается актуальной задачей. Так, в 2021–2022 гг. по данной теме опубликовано более 16 тыс. научных работ. Одна из основных причин смерти ковид-положительных пациентов — осложнения в работе сердечно-сосудистой системы (далее — ССС), вызванные воздействием коронавируса [1]. Для доклинической диагностики COVID-19 в основном используются два метода: биохимический на основе полимеразной цепной реакции (ПЦР-тест) и анализ крови. Необходимые в данном случае контакты с медперсоналом (в том числе визиты в медицинские учреждения) затрудняют регулярный оперативный контроль и повышают нагрузку на систему здравоохранения. Таким образом, представляется актуальным применение современных технологий доклинического контроля ССС для раннего выявления признаков COVID-19.

Регулярность контроля могут обеспечить носимые электронные устройства. Наиболее распространенные из них — фитнесс-браслеты и умные часы со встроенными датчиками частоты пульса и способностью выполнять измерения с высокой дискретностью [2]. Такой подход открывает возможности для анализа потоков данных на базе машинного обучения[1] [3].

Цель представленного исследования — создание обучаемой модели, способной выявлять ковидные аномалии, опираясь только на данные о ритме сердца. В ряде работ [4–6] рассматриваются подобные задачи, однако решения опираются на дополнительную информацию о состоянии крови и другие характеристики.[2] Это существенно ограничивает возможности их интеграции с носимыми устройствами, т. к. в моменте невозможно ввести в модель результаты анализа крови или мазка для ПЦР-теста. Новизна предложенного решения состоит в том, что использовались только данные ритма сердца, которые можно с высокой частотой снимать удобным для человека способом и интерпретировать показатели в режиме реального времени.

Материалы и методы

Характеристика данных. В работе использовали 229 обезличенных ритмограмм (кардиоинтервалографий) пациентов с COVID-19. Сведения получены в 2021 году в рамках открытого всероссийского соревнования для профессионалов в сфере цифровой экономики «Цифровой прорыв». Фрагмент данных представлен в таблице 1.

Таблица 1

Фрагмент набора данных

Номер ритмограммы	Время в миллисекундах	RR-интервал между сердцебиениями в миллисекундах	Признак ковидной аномалии*
81	0	576	0
81	568	568	0
81	1140	572	0
…	…	…	…

[1] Эндогенные аномалии кардиоритма у пациентов с COVID-19 / С. А. Пермяков [и др.] // Нелинейная динамика в когнитивных исследованиях — 2021 : тр. VII Всерос. конф. Нижний Новгород : Ин-т прикладной физики Российской академии наук, 2021. С. 109–110.
[2] Diagnosis of COVID-19 and its clinical spectrum / Kaggle Inc. // kaggle.com : [сайт]. URL: https://www.kaggle.com/datasets/einsteindata4u/covid19 (дата обращения : 10.09.2022).

Номер ритмограммы	Время в миллисекундах	RR-интервал между сердцебиениями в миллисекундах	Признак ковидной аномалии*
176	44332	568	0
176	44968	636	1
176	45596	628	0
*0 — аномалии нет, 1 — аномалия есть.			

На рис. 1 показана связь ритмограммы (RR interval) с электрокардиограммой сердца (ECG).

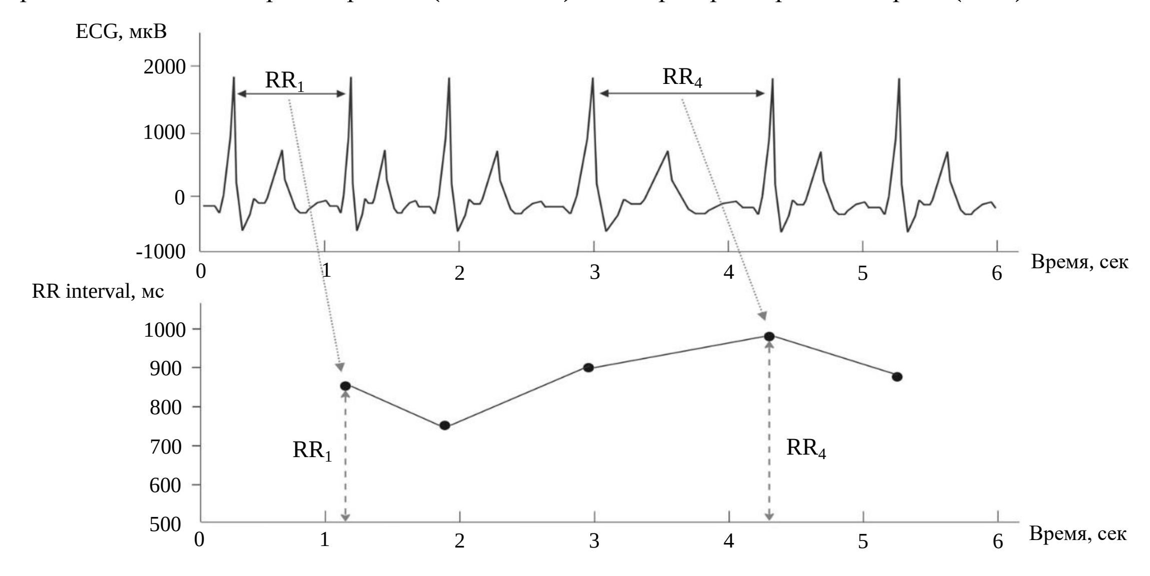


Рис. 1. Сопоставление электрокардиограммы и ритмограммы сердца: по горизонтальной оси показано время в секундах, по вертикальной для ЭКГ — микровольты

Во всех ритмограммах из этого набора есть промаркированные аномальные участки. На рис. 2 аномальные участки выделены красным пунктиром. По оси x показана продолжительность одного замера ритмограммы в миллисекундах, по y — интервал между соседними ударами сердца в миллисекундах.

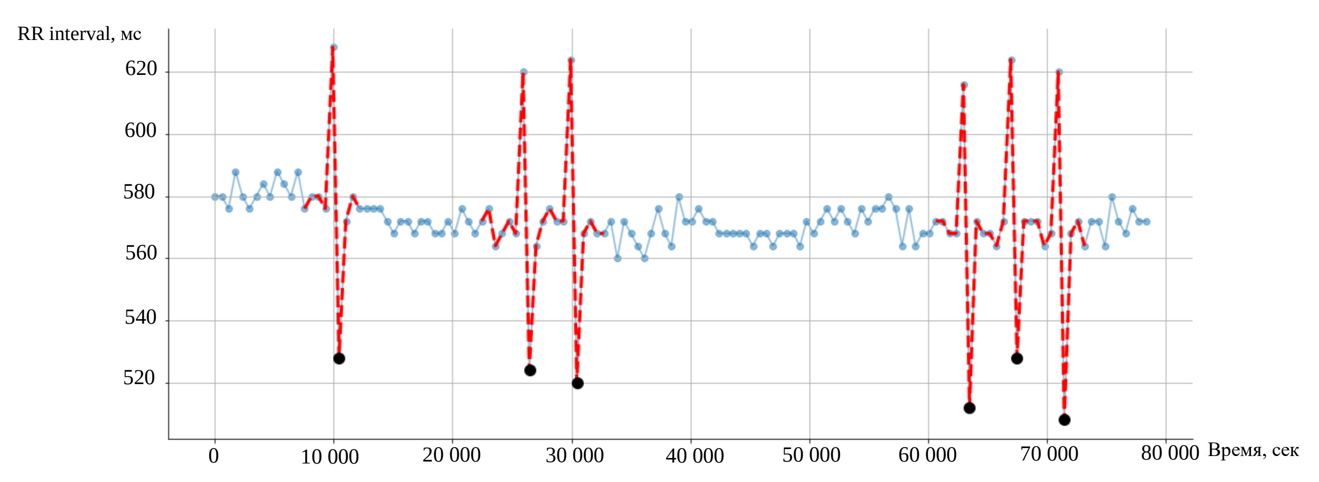


Рис. 2. График ритмограммы № 69: красным пунктиром выделены аномальные участки, черными буллитами — аномальные точки

Каждая ритмограмма представлена своим идентификатором. Продолжительность ритмограмм в исследуемом наборе данных различная: 4 минуты в среднем, 31 минута максимум. Каждый замер внутри одной ритмограммы имеет метку времени в миллисекундах от начала записи. Продолжительность RR-интервала также представлена в миллисекундах. Каждое конкретное значение в ритмограмме позволяет говорить о признаках аномалии (0 — нет, 1 — есть). 2,53 % наблюдений маркированы цифрой 1. Таким образом, набор данных имеет сильный дисбаланс классов, что типично для задач обнаружения аномалий.

В разметке данных встречаются различные подходы к выделению аномальных участков. Как аномальные выделялись группы точек в окрестности характерного пика и падения продолжительности ритма сердца: 3-го, 4-го, 6-го измерений (рис. 2). Не всегда количество точек в окрестности размечено одинаково — слева и справа от пика может быть разное количество аномальных точек. Кроме того, выявлены ритмограммы с

зашумленными показаниями. Так было при потере связи с гаджетом и замерах при установке или снятии прибора. 16 ритмограмм с некорректными данными пришлось исключить из рассмотрения, а разметку переделать:

– у аномального участка выделяется лишь одна точка, характеризующая аномальный фрагмент;

– аномальные точки обозначены черными буллитами (рис. 2).

Извлечение признаков. В чистом виде представлен лишь один сигнал — значение интервалов между сердцебиениями. Поэтому для уточнения модели подготовили дополнительные признаки на основании имеющегося сигнала: отклонение от медианного значения и производные в соседних замерах ритма. Этот перечень признаков выбрали после разведочного анализа данных и визуального выявления паттерна в местах, соответствующих аномальным участкам. На рис. 2 их обозначили красной пунктирной линией.

Результаты исследования

Метрика для оценки качества детекции аномалий. Для оценки качества модели в задаче бинарной классификации ввиду дисбаланса классов задействовали метрику F_1 [7] (1). Она позволяет оценить, насколько хорошо построенная модель детектирует редкий класс. В данном контексте под редким классом понимаются аномальные по продолжительности сердцебиения — сердцебиения с аномальным ритмом:

$$F_1 = 2 \times \frac{\text{точность} \times \text{полнота}}{(\text{точность} + \text{полнота})}. \quad (1)$$

Здесь:

– точность — доля правильно детектированных моделью аномальных сердцебиений от общего количества сердцебиений, которые модель определила как аномальные;

– полнота (или иными словами — чувствительность) — доля сердцебиений, которые модель верно детектировала как аномальные, от общего количества аномальных сердцебиений во всем наборе данных.

Алгоритмы машинного обучения. В рамках исследования применены пять описанных ниже алгоритмов машинного обучения.

1. IsolationForest — алгоритм с неконтролируемым самообучением на базе экстремально рандомизированных решающих деревьев [8].

2. Light Gradient Boosting Machine Classifier (LGBMClassifier) — алгоритм градиентного бустинга над решающими деревьями [9]. Для повышения скорости работы задействуются две техники: Gradient-based One-Side Sampling и Exclusive Feature Bundling[3].

3. RandomForestClassifier базируется на решающих деревьях и реализует многократный выбор случайного подмножества признаков. По ним строятся более простые оценщики — деревья решений. Результаты агрегируются для получения конечного предсказания [10].

4. ExtraTreesClassifier аналогичен RandomForestClassifier, однако в нем дополнительно реализован случайный выбор границы, по которой происходит ветвление узлов в деревьях решений [11].

5. SGDOneClassSVM[4] — линейная версия One-Class Support Vector Machine с использованием стохастического градиентного спуска.

IsoltionForest и SGDOneClassSVM были выбраны ввиду их широкого использования в задачах детекции аномалий [12, 13]. LGBMClassifier, RandomForestClassifier и ExtraTreesClassifier достаточно хорошо показывают себя в разных задачах, поэтому их тоже задействовали для сравнения результатов.

Особенность алгоритмов IsolationForest и SGDOneClassSVM заключается в том, что они не требуют на входе четкой разметки аномальных наблюдений, в то время как для остальных использованных в исследовании алгоритмов она обязательна.

IsolationForest базируется на предположении, что при построении изолирующих деревьев аномальные наблюдения можно изолировать (отделить) за меньшее количество операций, чем нормальные экземпляры наблюдений. Для каждого наблюдения алгоритм вычисляет значение оценки аномальности (anomaly score) по формуле:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}, \quad (2)$$

где $h(x)$ — число ребер до экземпляра x в каждом изолирующем дереве решений; $E(h(x))$ — среднее значение $h(x)$ на всем наборе изолирующих деревьев; $c(n)$ — нормализирующая константа для набора данных размером n (3).

[3] LightGBM: A Highly Efficient Gradient Boosting Decision Tree // www.microsoft.com : [сайт]. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/11/lightgbm.pdf (дата обращения: 10.09.2022).

[4] Online One-Class SVM / Scikit-learn developers (BSD License) // scikit-learn.org : [сайт]. URL: https://scikit-learn.org/stable/modules/sgd.html#online-one-class-svm (дата обращения: 10.09.2022).

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}, \quad (3)$$

$$H(k) = \ln(k) + \gamma. \quad (4)$$

В уравнении (4) γ — постоянная Эйлера, равная 0,57721…

Если наблюдение x имеет значение оценки аномальности s, близкое к 1, то оно считается аномальным. Если s близко к 0,5, то наблюдение не имеет очевидных признаков аномальности. Если s близко к 0, то наблюдение может считаться нормальным (рис. 3).

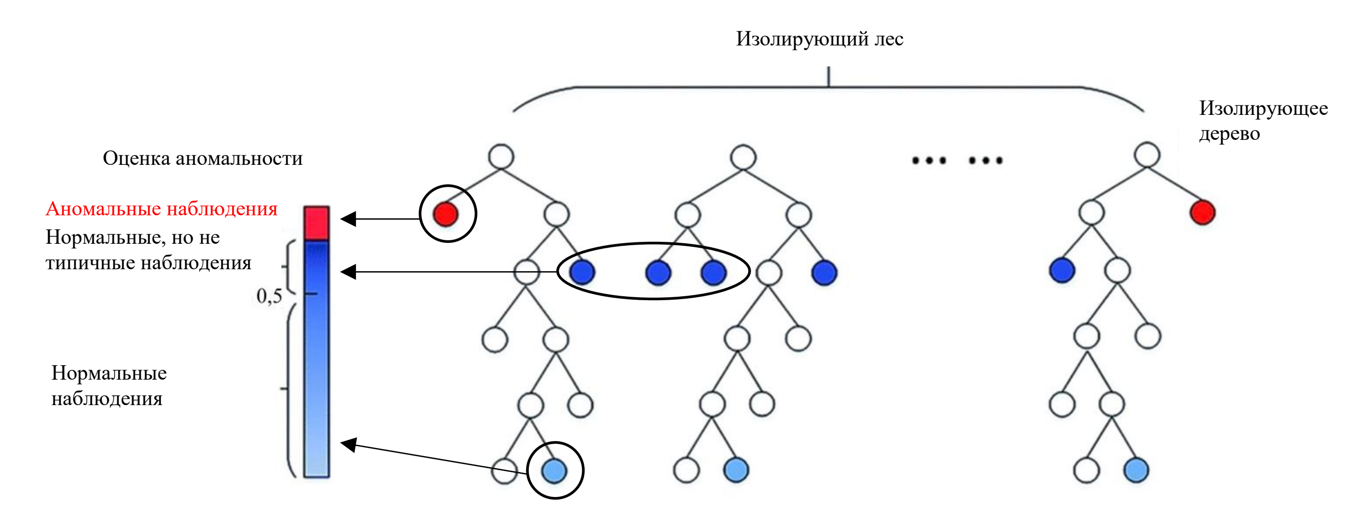


Рис. 3. Нормальные и аномальные наблюдения в изолирующих деревьях

SGDOneClassSVM основывается на противоположном относительно IsolationForest подходе. Алгоритм определяет границы нормальных наблюдений и все новые наблюдения сопоставляет с границами этой нормы, чтобы выявить аномалию.

Значимость признаков. Оценка степени влияния признаков на прогностическую способность модели представлена на рис. 4.

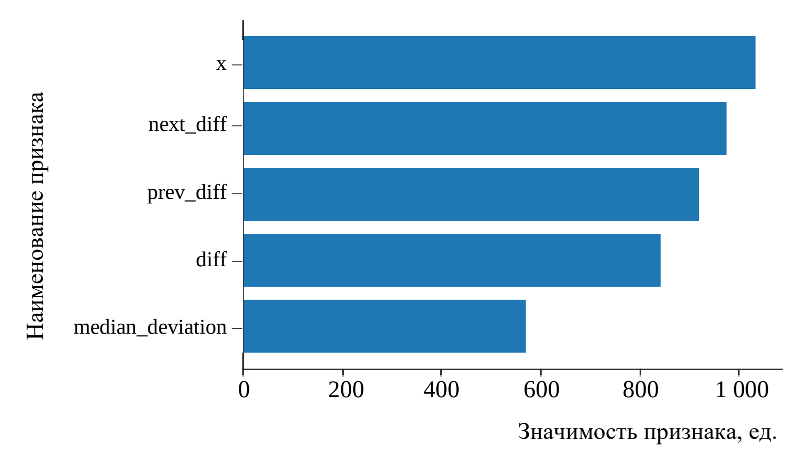


Рис. 4. Диаграмма значимости признаков: x — текущее значение интервала; next_diff — производная в последующей точке измерения продолжительности сердцебиения; prev_diff — производная в предыдущей точке измерения продолжительности сердцебиения; diff — первая производная в текущей точке; median_deviation — отклонение текущего значения длительности *RR*-интервала от медианы

Для расчета числовой оценки значимости использовался встроенный в LGBMClassifier механизм, возвращающий через свойство *feature_importances_* обученной модели массив числовых оценок для каждого признака. Значимость в моделях на основе градиентного бустинга над решающими деревьями, как правило, рассчитывается на основе индекса Джини (Gini-impurity Index[5]) [14], используемого в процессе определения точек ветвления при обучении модели:

$$Gini(d) = 1 - \sum_{i=1}^{k} p_i^2 \,. \quad (5)$$

[5] Karabiber F. Gini Impurity // learndatasci.com : [сайт]. URL: https://www.learndatasci.com/glossary/gini-impurity/ (дата обращения: 10.09.2022).

Здесь d — набор наблюдений, подходящих по условиям в рассматриваемой точке ветвления, $d \in D$; k — количество классов, представленных во всем тренировочном наборе данных D; p_i — вероятность принадлежности наблюдений к классу i в рассматриваемой точке ветвления решающего дерева.

Самыми значимыми оказались следующие признаки: текущее значение интервала (x), производная в последующей (next_diff) и предыдущей (prev_diff) точках измерения продолжительности сердцебиения (рис. 4). Полный перечень используемых признаков приведен в таблице 2:

Таблица 2

Перечень используемых признаков

№	Признак	Описание
1	*x*	*RR*-интервал в текущей точке измерения
2	next_diff	Первая производная в следующей точке
3	prev_diff	Первая производная в предыдущей точке
4	diff	Первая производная в текущей точке
5	median_deviation	Отклонение текущего значения длительности *RR*-интервала от медианы в рамках одной ритмограммы

Сравнение моделей. Для результативности модели SGDOneClassSVM важен подбор параметра *nu*, который соответствует вероятности обнаружения регулярного наблюдения за пределами границы нормы. Иными словами, *nu* определяет верхнюю границу доли ошибок при обучении модели и нижнюю границу доли опорных векторов.[6] Для подбора *nu* с учетом природы имеющихся данных дополнительно оценивалась метрика качества при различных значениях указанного параметра (рис. 5). В итоге выбран *nu*, равный 0,11.

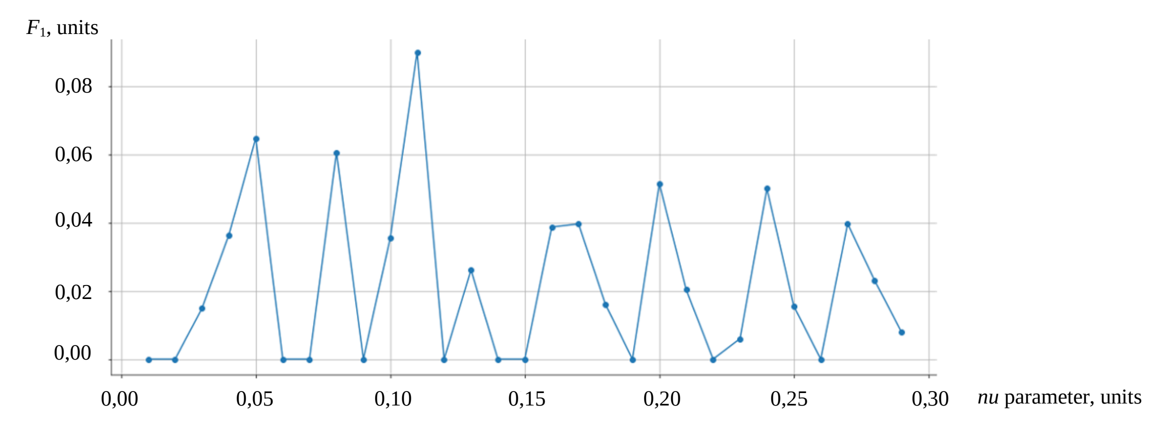


Рис. 5. Оценка параметра *nu* (по горизонтальной оси) для модели *SGDOneClassSVM*. На вертикальной оси — значения метрики F_1

Для расчета метрики качества на различных моделях использовался весь набор данных с применением техники перекрестной проверки. Внутри одной ритмограммы мы имеем временной ряд наблюдений, снятых за один непрерывный промежуток времени у одного человека, поэтому следует рассматривать их как зависимые [15]. Для разделения данных на обучающие и тестовые наборы применялась следующая стратегия. Отобранный набор данных состоит из 213 ритмограмм, помеченных уникальным идентификатором (*id*). Это дает возможность выделить ритмограммы для обучения и тестирования моделей. Набор ритмограмм для теста можно случайным образом выбирать по идентификаторам. Ниже описан подход, примененный в представленной работе.

I. В цикле разделения данных выполняются пять действий.

1. Фиксируется начальное число для генерации псевдослучайных чисел (seed) —np.random.seed(fold), где fold — номер текущего разбиения данных.

2. Генерируется 42 случайных целочисленных значения в диапазоне от 1 до 213. Так мы получаем случайные номера идентификаторов ритмограмм для тестового набора данных.

[6] SGDOneClassSVM documentation. Scikit-learn developers (BSD License) scikit-learn.org [сайт]. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDOneClassSVM.html#sklearn.linear_model.SGDOneClassSVM (дата обращения: 10.09.2022).

3. Вносятся в отдельный список номера идентификаторов ритмограмм, которые остались после отбора идентификаторов для теста. Их задействуют для тренировочного набора.

4. На ритмограммах из тренировочного набора обучаются модели, а на ритмограммах из тестового набора оцениваются метрики качества прогнозирования.

5. Записывается значение метрики качества для каждой модели, посчитанной на тестовом наборе ритмограмм при текущем разбиении данных.

II. Шаги 1–5 повторяются для каждого номера разбиения данных.

III. Полученные значения метрики качества усредняются для каждой из моделей.

Сравнительная оценка усредненной метрики качества прогнозирования для каждой модели приведена в таблице 3.

Таблица 3

Оценка метрики качества F_1

Модель	Метрика F_1*
LGBMClassifier	0,8328
RandomForestClassifier	0,7638
ExtraTreesClassifier	0,7369
SGDOneClassSVM	0,0169
IsolationForest	< 1e-4
*Среднее значение при выбранной стратегии перекрестной проверки на пяти разбиениях.	

Обсуждение и заключения. Разработана математическая модель обнаружения аномалий в ритме сердца с точностью 83 %. По метрике качества F_1 лучшей оказалась модель на базе алгоритма LGBMClassifier. IsolationForest и SGDOneClassSVM на текущих данных показали слабые результаты.

Предложенную модель можно реализовать в составе программной части носимых персональных смарт-устройств. Предлагаемый сценарий использования решения:

– запись ритмограммы активируется на персональном носимом устройстве через пользовательский интерфейс;

– по окончании запись подается в разработанную модель для анализа;

– по итогам анализа данных математическая модель выдает уведомление о наличии или отсутствии аномалий на экране носимого устройства.

Отметим, что для записи одной ритмограммы, видимо, достаточно в среднем 4 минут. За это время возможно обнаружение ковидных аномалий в ритме сердца.

Модель занимает в памяти носимого устройства 493 килобайтов, что вполне подходит для практического применения. Решение опирается только на информацию о ритме сердца и не задействует факторы, недоступные для мобильных персональных гаджетов.

Повышение точности детекции аномалий предполагает дополнительные изыскания. Их следует сфокусировать на разработке уникальных признаков, которые выявляются по исходному сигналу ритма сердца. Однако текущее решение уже дает возможность оперативной и легкой оценки вероятности COVID-19 на ранней стадии. Это наряду с выполнением рекомендаций медиков может дополнительно способствовать снижению риска смертности от негативного влияния коронавирусной инфекции на сердечно-сосудистую систему.

Список литературы

1. Турсунова Н.Д., Шафигулина И.С., Гребенникова И.В. и др. Патогенетические аспекты влияния COVID-19 на сердечно-сосудистую систему человека. *European Journal of Natural History*. 2022;1:73–77.

2. Молодченков А.И., Григорьев О.Г., Шарафутдинов Я.Н. Автоматическое выявление значений факторов риска заболеваний с помощью методов искусственного интеллекта и технологии интернета вещей. *Информационные технологии и вычислительные системы*. 2021;1:83–96. https://doi.org/10.14357/20718632210109

3. Polevaya S.A., Eremin E.V., Bulanov N.A., et al. Event-Related Telemetry of Heart Rhythm for Personalized Remote Monitoring of Cognitive Functions and Stress under Conditions of Everyday Activity. *Modern Technologies in Medicine*. 2019;11(1):109–115. http://dx.doi.org/10.17691/stm2019.11.1.13

4. Kouame Amos Brou, Ivan Smirnov, Mabouh Moise Hermann. Comparison of Machine Learning Models for Coronavirus Prediction. *Advanced Engineering Research (Russia).* 2022;22(1):67–75. https://doi.org/10.23947/2687-1653-2022-22-1-67-75

5. Ashish Bhargava, Elisa Akagi Fukushima, Miriam Levine, et al. Predictors for Severe COVID-19 Infection. *Clinical Infectious Diseases.* 2020;71:1962–1968. https://doi.org/10.1093/cid/ciaa674

6. Красюкова Ю.И., Вахрушева Т.А. Модель машинного обучения для определения вероятности заболевания COVID-19 по первичным признакам. *Интеллектуальные ресурсы — региональному развитию.* 2021;2:67–71.

7. Alaa Tharwat. Classification Assessment Methods. *Applied Computing and Informatics*. 2021;17(1):174.https://doi.org/10.1016/j.aci.2018.08.003

8. Yupeng Xu, Hao Dong, Mingzhu Zhou, et al. Improved Isolation Forest Algorithm for Anomaly Test Data Detection. *Journal of Computer and Communications.* 2021;9(8):49–51. https://doi.org/10.4236/jcc.2021.98004

9. Bruce P., Bruce A., Gedeck P. *Practical Statistics for Data Scientists*, 2nd ed. Boston: O'Reilly Media; 2020. 342 p.

10. Breiman L. Random Forests. *Machine Learning.* 2001;45:5–32. https://doi.org/10.1023/A:1010933404324

11. Geurts P., Ernst D., Wehenkel L. Extremely Randomized Trees. *Machine Learning.* 2006;63:3–42. https://doi.org/10.1007/s10994-006-6226-1

12. Kaur H., Singh G., Minhas J. A Review of Machine Learning Based Anomaly Detection Techniques. *International Journal of Computer Applications Technology and Research*. 2013;2(2):185–187. http://dx.doi.org/10.7753/IJCATR0202.1020

13. Кацер Ю.Д., Козицин В.О., Максимов И.В. Методы обнаружения неисправностей оборудования АЭС. *Известия высших учебных заведений. Ядерная энергетика.* 2019;4:5–27. https://doi.org/10.26583/npe.2019.4.01

14. Daniya T., Geetha M., Suresh Kumar K. Dr. Classification and Regression Trees with Gini Index. *Advances in Mathematics Scientific Journal.* 2020;9(10):8237–8247. http://dx.doi.org/10.37418/amsj.9.10.53

15. Valliappa Lakshmanan, Sara Robinson, Michael Munn. *Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps*, 1st ed. Boston: O'Reilly Media; 2020. 408 p.

References

1. Tursunova ND, Shafigulina IS, Grebennikova IV, et al. Patogeneticheskie aspekty vliyaniya COVID-19 na serdechno-sosudistuyu sistemu cheloveka. *European Journal of Natural History.* 2022;1:73–77. (In Russ.)

2. Molodchenkov AI, Grigoriev OG, Sharafutdinov YaN. Automatic Calculation of Disease Risk Factors Values Using Artificial Intelligence Methods and Internet of Things Technology. *Journal of Information Technologies and Computing Systems.* 2021;1:83–96. https://doi.org/10.14357/20718632210109

3. Polevaya SA, Eremin EV, Bulanov NA, et al. Event-Related Telemetry of Heart Rhythm for Personalized Remote Monitoring of Cognitive Functions and Stress under Conditions of Everyday Activity. *Modern Technologies in Medicine*. 2019;11:109–115. http://dx.doi.org/10.17691/stm2019.11.1.13

4. Kouame Amos Brou, Ivan Smirnov, Mabouh Moise Hermann. Comparison of Machine Learning Models for Coronavirus Prediction. *Advanced Engineering Research (Russia).* 2022;22:67–75. https://doi.org/10.23947/2687-1653-2022-22-1-67-75

5. Ashish Bhargava, Elisa Akagi Fukushima, Miriam Levine, et al. Predictors for Severe COVID-19 Infection. *Clinical Infectious Diseases.* 2020;71:1962–1968. https://doi.org/10.1093/cid/ciaa674

6. Krasyukova YuI, Vakhrusheva TA, Pei He Su. Machine Learning Model for Determining the Probability of Covid-19 Disease by Primary Signs. *Intellektual'nye resursy — regional'nomu razvitiyu*. 2021;2:67–71.

7. Alaa Tharwat. Classification Assessment Methods. *Applied Computing and Informatics.* 2021;17:174. https://doi.org/10.1016/j.aci.2018.08.003

8. Yupeng Xu, Hao Dong, Mingzhu Zhou, et al. Improved Isolation Forest Algorithm for Anomaly Test Data Detection. *Journal of Computer and Communications.* 2021;9:49–51. https://doi.org/10.4236/jcc.2021.98004

9. Bruce P, Bruce A, Gedeck P. *Practical Statistics for Data Scientists*, 2nd ed. Boston: O'Reilly Media; 2020. 342 p.

10. Breiman L. Random Forests. *Machine Learning.* 2001;45:5–32. https://doi.org/10.1023/A:1010933404324

11. Geurts P, Ernst D, Wehenkel L Extremely Randomized Trees. *Machine Learning.* 2006;63:3–42. https://doi.org/10.1007/s10994-006-6226-1

12. Kaur H, Singh G, Minhas J. A Review of Machine Learning Based Anomaly Detection Techniques. *International Journal of Computer Applications Technology and Research.* 2013;2:185–187. http://dx.doi.org/10.7753/IJCATR0202.1020

13. Katser ID, Kozitsin VO, Maksimov IV. NPP Equipment Fault Detection Methods. *Proc. of Universities. Nuclear Power Engineering*. 2019;4:5–27. https://doi.org/10.26583/npe.2019.4.01

14. Daniya T, Geetha M, Suresh Kumar K Dr. Classification and Regression Trees with Gini Index. *Advances in Mathematics Scientific Journal*. 2020;9:8237–8247. http://dx.doi.org/10.37418/amsj.9.10.53

15. Valliappa Lakshmanan, Sara Robinson, Michael Munn. *Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps,* 1st ed. Boston: O'Reilly Media; 2020. 408 p.

Об авторах:

Межов Максим Сергеевич, ведущий эксперт ООО «Цифровые технологии и платформы» (115054, РФ, Москва, ул. Дубининская, 53, стр. 6), ORCID, msmezhov@ya.ru

Козицин Вячеслав Олегович, ведущий эксперт ООО «Цифровые технологии и платформы» (115054, РФ, Москва, ул. Дубининская, 53, стр. 6), ORCID, Vyacheslav.Kozitsin@skoltech.ru

Кацер Юрий Дмитриевич, аспирант сколковского института науки и технологии (121205, РФ, Москва, территория инновационного центра «Сколково», Большой бульвар, 30, стр. 1), ScopusID, ORCID, Iurii.katser@skoltech.ru

Заявленный вклад соавторов:

М.С. Межов — формирование основной концепции, цели и задач исследования, сбор данных, разработка моделей, расчеты и анализ результатов. В.О. Козицин — подготовка текста, формулирование выводов, предварительная обработка данных и доработка текста. Ю.Д. Кацер — контроль проведения исследования, доработка текста и корректировка выводов.

Поступила в редакцию 09.12.2022.
Поступила после рецензирования 25.01.2023.
Принята к публикации 25.01.2023.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the Authors:

Maksim S Mezhov, leading expert, "Digital Technologies and Platforms" LLC (53, Dubininskaya St., Moscow, 115054, RF), ORCID, msmezhov@ya.ru

Vyacheslav O Kozitsin, leading expert, "Digital Technologies and Platforms" LLC (53, Dubininskaya St., Moscow, 115054, RF), ORCID, Vyacheslav.Kozitsin@skoltech.ru

Iurii D Katser, postgraduate, Skolkovo Institute of Science and Technology (30, Bolshoy Boulevard, Moscow, 121205, RF), ScopusID, ORCID, Iurii.katser@skoltech.ru

Claimed contributorship:

MS Mezhov: basic concept formulation; research objectives and tasks; data collection; model development; calculations and analysis of the results. VO Kozitsin: text preparation; formulation of conclusions; pre-processing of data; the text revision. ID Katser: control of the study; revision of the text; correction of the conclusions.

Received 09.12.2022.
Revised 25.01.2023.
Accepted 25.01.2023.

Conflict of interest statement

The authors do not have any conflict of interest.

All authors have read and approved the final manuscript.